(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平7－222129

(43)公開日　平成７年(1995)８月18日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｈ０４Ｎ　7/15				

審査請求　未請求　請求項の数９　ＯＬ　（全 6 頁）

(21)出願番号　特願平6－8133

(22)出願日　平成６年(1994)１月28日

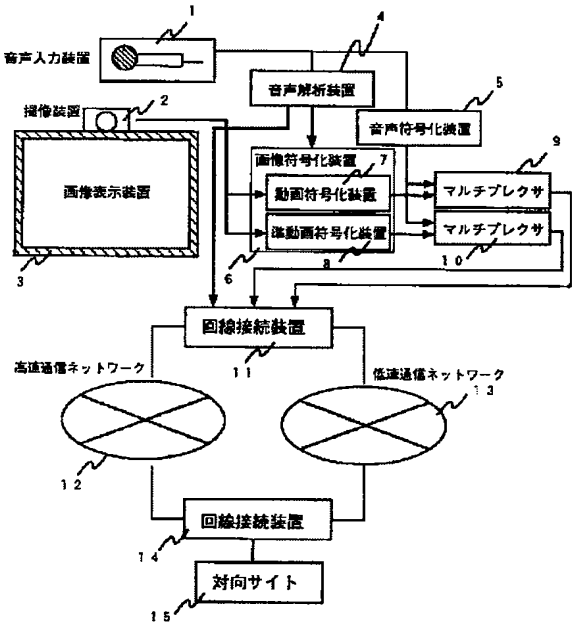
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　三村　到
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　鈴木　敏明
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　亀山　達也
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内
(74)代理人　弁理士　小川　勝男

最終頁に続く

(54)【発明の名称】　映像通信装置

(57)【要約】
【構成】音声の中から動画モードと準動画モードの切替を指示する言葉を検出する手段と、映像の中から動画モードと準動画モードの切替を指示するジェスチャを検出する手段とを備え、検出した指示に応じて符号生成量を動画（大量）／準動画（小量）に切替え可能な符号化装置６を備える。
【効果】会話を必要としない映像接続では準動画の伝送のため通信コストの節約が図れ、会話が必要な場合には高精細動画を用いた臨場感映像通信が可能となる。

図１



1

【特許請求の範囲】

【請求項１】映像入力手段と、映像表示手段と、映像符号化伝送手段とを含む映像通信装置をネットワークで接続して映像通信を行う装置において、音声入力手段と、前記音声入力手段により入力した音声信号を解析する手段と、前記音声信号の解析手段によって制御される映像符号化手段を備えたことを特徴とする映像通信装置。

【請求項２】映像入力手段と、映像表示手段と、映像符号化伝送手段とを含む映像通信装置をネットワークに接続して映像通信を行う装置において、画像解析手段と、前記画像解析手段により解析した画像信号の特徴に応じて符号生成量が制御される画像符号化手段とを備えたことを特徴とする映像通信装置。

【請求項３】映像入力手段と、映像表示手段と、映像符号化伝送手段とを含む映像通信装置をネットワークに接続して映像通信を行う装置において、音声解析手段と、画像解析手段と、前記音声解析手段と前記画像解析手段とによって符号生成量を制御される画像符号化手段とを備えたことを特徴とする映像通信装置。

【請求項４】請求項１において、前記音声信号を解析する手段が認識する音声信号は、「もしもし」，「おーい」，「すみません」，「誰かいませんか」，「応答して下さい」，「やあ」，「さよなら」，「ばいばい」，「ではまた」，「じゃー」からなる呼掛けに使用する言葉である映像通信装置。

【請求項５】請求項１において、前記音声信号を解析する手段が認識する音声信号は、前記映像通信装置の使用者の姓、もしくは名、もしくは姓名、もしくは前記姓名にさん付けしたもの、もしくは役職名を付加したものである映像通信装置。

【請求項６】請求項１において、前記映像信号を解析する手段が解析する画像信号が人間の動作である映像通信装置。

【請求項７】請求項６において、前記人間の動作が「手を振る」，「指を指す」，「画面を見つめる」，「背を向ける」である映像通信装置。

【請求項８】請求項１において、解析すべき前記音声信号を登録可能な映像通信装置。

【請求項９】請求項２において、解析すべき前記映像信号の特徴を登録可能な映像通信装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、映像によりサテライトオフィス間を接続する映像通信装置に関する。

【０００２】

【従来の技術】従来より映像通信を利用して会議を行うテレビ会議装置が実用に供している。テレビ会議装置では、カメラで撮影した映像信号をデジタル化し、さらにこの信号を圧縮して伝送している。このように画像を圧縮する理由は以下の通りである。すなわち、テレビ解像

2

度の映像を完全な動画で伝送するには、１００メガビット／秒を越える高速な映像伝送ネットワークを使用する必要があり、このような高速回線の伝送コストは極めて高価なためである。そこで、画像信号の持つ統計的特徴を利用して冗長性を取り除く等の手法により１００メガビット／秒以上の映像データを、数十キロビット／秒から数メガビット／秒に圧縮して伝送する手法が開発されてきた。現在では、ＮＴＴのサービスするＩＳＤＮ回線（Integrated Services on Digital Network：６４キロビット回線）に収容できるほど高効率に圧縮できる技術が開発されてきている。しかし、今までに開発されてきた画像圧縮手法といえども、回線容量内に符号量発生量を抑えるため、駒落しや解像度低下が生じ、臨場感ある映像通信は必ずしも可能とは言えない。また、最近では、さらに伝送容量を増したＢＩＳＤＮ（Broadband ISDN）と呼ばれる通信サービスも始まる予定ではあるが、この回線を使用するとしても依然として回線使用量は高く、映像通信サービスを普及させる障害となっている。

【０００３】

【発明が解決しようとする課題】ところで、首都圏の地価高騰を背景に事務所やオフィスを郊外に設置するような形態（以後、サテライトオフィスと称する）が出現してきている。このサテライトオフィスでは、電話，ファックス、またはコンピュータネットワークを使用してオフィス間を接続し、物理的な距離を感じさせずに作業が可能な環境を提供しようとするもので、１９９３年時点で幾つかの試行がある。これらの試行を通じてサテライトオフィスをより実用的なものとするには、上記に示した通信装置だけでなく、お互いのオフィスの映像を常時表示して同一空間の共有感を創造する映像通信装置を利用することが効果的であると指摘されている。この理由は、相手地点の映像を常時表示することで同じ空間を共有しているといった感覚を生じさせ、これにより遠隔地のオフィスにいるといった疎外感を軽減することの効果が考えられている。

【０００４】ところが、上述のような常時接続する映像通信に、これまでに普及しているテレビ会議装置や映像伝送装置、さらにこれらを接続する回線としてＩＳＤＮ回線やＢＩＳＤＮ回線を使用すると、一定量の符号を常に発生するため通信回線を常に占有することになる。従って通信コストが高額になるといった問題がある。

【０００５】

【課題を解決するための手段】そこで本発明では、通信コストの上昇を抑制しつつ作業効率の向上が図れる臨場感映像通信装置を提供するため、会話の開始と終了を検出するための音声信号解析手段と、画像符号化特性を動画と準動画に切替えられる映像符号化手段と、画像符号化方式に応じて接続する回線を選択する回線接続手段とを設ける。

【０００６】

【作用】上記に示した画像符号化手段は、符号発生量は多いが高精細で動きの滑らかな動画信号を発生するモードと、動きに対しては追従性は劣るが符号発生量の少ない符号化モードを備える。音声解析手段は、ユーザの発する音声信号の内容を理解して、必要に応じて符号化特性を切替える。例えば、映像表示画面の前に位置するユーザが臨場感のある会話を行う場合、すなわち、高精細動画モードを必要とする場合は、そのユーザが発生する音声として、「もしもし」，「おーい」などといった言葉を検出し、動画対応の符号化を行うように符号化装置を制御する。この認識装置は、同時に使用する回線を選択し、完全動画モードでは発生する大量のデータを伝送可能な回線に接続するように動作する。一方、会話を終了する際は、会話終了のキーとなる語を検出し、符号化特性を準動画（符号発生量の少ないモード）となるように制御する。これにより、会話時は動きの滑らかで臨場感の高い映像通信が可能となり、また会話のない時は低コストの映像通信が行える。

【０００７】

【実施例】図１は本発明による映像通信装置の構成を示したブロック図である。本装置は、音声入力装置１，撮像装置２，画像表示装置３，音声解析装置４，音声符号化装置５，画像符号化装置６，マルチプレクサ９，１０、および回線接続装置１１から構成される。この映像通信装置は、通信ネットワークを介して遠方にある他の通信装置と接続されている。なお本発明の装置は大容量と少容量のデータを発生する画像符号化装置を備えているため、図１の実施例ではそれぞれの容量に対応した二系統のネットワークを使用することを前提に図を描いてある。もちろん回線容量が可変なネットワークでは、一つの回線に異なるデータ量をもつ信号を接続できるので、二つの異なるネットワークに接続する必要がないことは容易に理解できる。なお、この二系統のデータ発生方法については後述する。

【０００８】次に図１の装置の動作を説明する。図１の映像通信装置は、撮像装置２で得た映像信号を画像符号化装置６により符号化・圧縮する。映像と同時に収録された音声信号は音声符号化装置５により符号化し、マルチプレクサ９，１０により映像信号と合成する。音声信号は、同時に音声解析装置４にも入力しておく。なおこの音声解析装置の動作についても後述する。

【０００９】回線接続装置１１はこの映像通信装置をネットワークに接続するもので、先に述べた映像・音声信号はネットワークを介して相手の映像通信装置へと送出される。図１の実施例が、従来の映像通信装置と異なる点は、画像符号化装置が二種類の異なるデータレートの画像符号化装置を有し、発生するデータ量を状況に応じて制御する点である。データレートの異なる画像符号化装置として、ここでは動画符号化装置７と準動画符号化装置８を含む構成として記してある。すなわち、動画符号化装置７は、駒落しすることなく動きの滑らかな映像を再生するための符号化装置である。従って動画符号化装置は、毎秒に符号化する映像フレーム数が多くなる。そのため、必然的に生成データ量は多くなる。一方、準動画符号化装置８は、駒落しを行うことでデータ発生量の少ない画像符号化を行う装置である。この二系統の符号化装置のいずれを使用するかの制御は音声解析装置４が行う。この音声解析装置４は、同時に回線接続装置１１も制御し、この二種類の装置を制御することで、回線に送出する映像・音声データの量を制御する。

【００１０】次に、この音声解析装置の動作を詳細に説明する。音声解析装置４は、音声入力装置１からの音声信号を解析し、会話の開始／終了を検出する。会話を開始するなど、臨場感の高い動画符号化に対する要求がある場合は、動画像符号化装置７により符号化を行い、また要求が無い場合は準動画による符号化を行うように画像符号化装置６を制御する。またこの時、同時に回線接続装置１１をも制御し、発生符号量に応じた回線に接続するように動作する。

【００１１】図２は音声解析装置４の構成を詳細に示したブロック図である。本音声解析装置は音声照合装置２１と登録語データベース２２から構成する。この登録語データベース２２には動画モードを開始するためのキーワードを登録した動画開始予約語データベース２２（図中には動画開始予約語と記載）と動画モードを終了するためのキーワードを登録した動画終了予約語データベース２３（図中には動画終了予約語と記載）を備えている。

【００１２】信号入力端子２０に入力された音声信号は音声照合装置２１に入力する。この音声信号は音声認識によりその言葉を解釈する。この言葉（以後、語と記す）は、動画開始予約語データベース２３の予約語、動画終了予約語データベースの予約語と照合する。

【００１３】動画開始の予約語は、例えば「もしもし」，「おーい」といった呼掛けの言葉や、その他、システムを使用する人名や、その人名に「さん」や肩書を付加したものである。すなわち、会話を開始するために通常良く使用する言葉である。このような語を動画モードの開始予約語として登録しておき、この語を検出した時点で会話状態フラグを会話に設定し、動画による符号化を開始させる。

【００１４】一方、動画終了予約語は、例えば、「さよなら」，「じゃー」，「では」，「ばいばい」といった通常会話を終りにする際に使用する語である。このような会話を終了する時に使用する語を登録しておき、この語により会話終了を検出することで動画符合化モードを終了して準動画モードへと移行するようなフラグ信号を発生する。

【００１５】このように会話の開始語と、会話の終了語を用いて符号化モードを切替えることの効果を次に説明

する。通常、オフィスを映像で結んだ際には、会話が発生していない時に動画を表示することと、駒数を減じた準動画を表示することの差は少なく、動画を伝送しておくことは回線コストがかかるが得られる効果は少ない。従って、会話の無い時は準動画の伝送で十分である。これに対し、通信相手と会話を行おうとする時に駒数の少ない準動画では、動きがぎこちなくなり自然な会話が行えない。そこでこの時は、駒数の豊富な動画信号を使用するのが好ましい。

【００１６】音声解析により得られる効果は、会話の開始語と終了語の間にある期間を会話期間と解釈し、この間は高画質な動画符号化により臨場感ある会話環境をサポートし、それ以外の期間は低コストな映像伝送に切替える制御を可能とする。

【００１７】図３は画像符号化装置の変形例を表す図である。図１の実施例では、動画符号化装置７と準動画符号化装置８の二系統の符号化装置を用い、これらを音声解析の結果に応じて切替えて使用する構成であったが、図３の変形例では、可変符号化装置により動的に符号発生量を制御するような構成にしてある。この構成によれば、符号化器を二系統持つ必要がなくなり、回路規模の削減が図れる。この可変符号化装置は、画像符号化装置３１、符号化レート制御装置３２、レート監視装置３３から構成する。レート監視装置３３は、音声解析装置４の解析結果に応じて、発生する符号量を決定する。先にも述べたように音声解析装置４は符号生成量を制御するが、それには、レート監視装置３３に対して会話開始／終了を指示する会話状態フラグ入力して制御する。

【００１８】図４は、図１の実施例の変形を表す図である。図１の実施例と図４の実施例の大きく異なる点は画像解析装置４０を設けた点である。この画像解析装置４０によって画像符号化装置４３、および回線接続装置４１を制御する点が新規な点である。図３の画像解析装置４０を設けた目的は、音声認識装置４を用いるのと同様であり、会話の開始／終了を検出するためである。そのため、画像解析装置は撮像した映像の中から、会話の開始を意味する動作（以下、ジェスチャと記す）を検出し、このジェスチャに基づいて画像符号化装置４３と回線接続装置４１を制御する。制御方法は先ほどの音声解析の場合と同様に、会話期間は高画質の動画を伝送するように、また会話の終了以後は符号量の少ない準動画を伝送するように制御する。

【００１９】図５は画像解析装置４０の構成を示した実施例である。画像解析装置４０は画像照合装置５２とジェスチャデータベース５２から構成する。このジェスチャデータベース５２は動画開始予約データベース５３と動画終了データベース５４からなる。この動画開始予約データベースには、会話の開始を意味するジェスチャとして、「手を振る」，「画面を見つめる」，「指で指す」，「手招きをする」などの動作を登録しておく。こ

のように登録したジェスチャにより会話の開始を検出し、会話が始まった場合には会話の開始を表すフラグ信号を発生する。また、動画終了予約ジェスチャには、「手を振る」，「背を向ける」などのジェスチャを登録しておき、このようなジェスチャを認識した時に会話の終了を表す終了フラグを発生する。

【００２０】なお、ジェスチャの照合方式は本発明の趣旨とは直接関係ないので、ここでの詳細な説明は省略するが、例えば「手を振る」動作を認識するには、手の形のテンプレートをデータベースに用意しておき、このテンプレートとのパターンマッチングにより手を認識し、さらにこの手の動きベクトルを検査することで、「手を振る」といった動作が解析できる。なお、この説明はジェスチャ認識のほんの一例であり、どのような認識方法であっても本発明の趣旨が満足されることは容易に理解できる。

【００２１】また、図５の実施例では、会話の開始と終了と共に「手を振る」動作が含まれているが、これはトグル操作の意味であり、始めに「手を振る」動作が会話の開始、次の「手を振る」動作が会話の終了といったように解釈すればよい。

【００２２】さらに、本装置による使い勝手を向上させるために、図２の音声解析装置のキーワードや、図５の画像解析装置のキーとなるジェスチャはユーザが登録可能な構成にしておくことが好ましい。すなわち、通常良く使用する言葉やジェスチャを必要に応じて登録することで、会話の開始／終了の照合をより確実とすることが可能となる。

【００２３】本実施例によれば音声信号、あるいは表示画面の前に立った人物のジェスチャを解析することで会話の始めと終りを検出し、会話が行われている期間だけ画像符号化装置を高画質な動画モードで動作させる。一方、会話の行われていない期間は、駒数を減じてデータ発生量の少ない準動画符号化とする。

【００２４】

【発明の効果】本発明によれば、ユーザが画面に写った相手オフィスの映像に呼掛けを行なった時だけコストのかかる伝送路を使用し、その他の場合には通信コストの低い回線を使用する。その結果として会話の時の臨場感を確保しつつ、通信コストの少ない映像通信が可能となる。

【図面の簡単な説明】

【図１】本発明による映像通信装置のブロック図。

【図２】音声解析装置のブロック図。

【図３】映像符合化装置の変形例のブロック図。

【図４】ジェスチャ解析装置を含む映像通信装置のブロック図。

【図５】ジェスチャ解析装置のブロック図。

【符号の説明】

１…音声入力装置、２…撮像装置、３…表示装置、４…

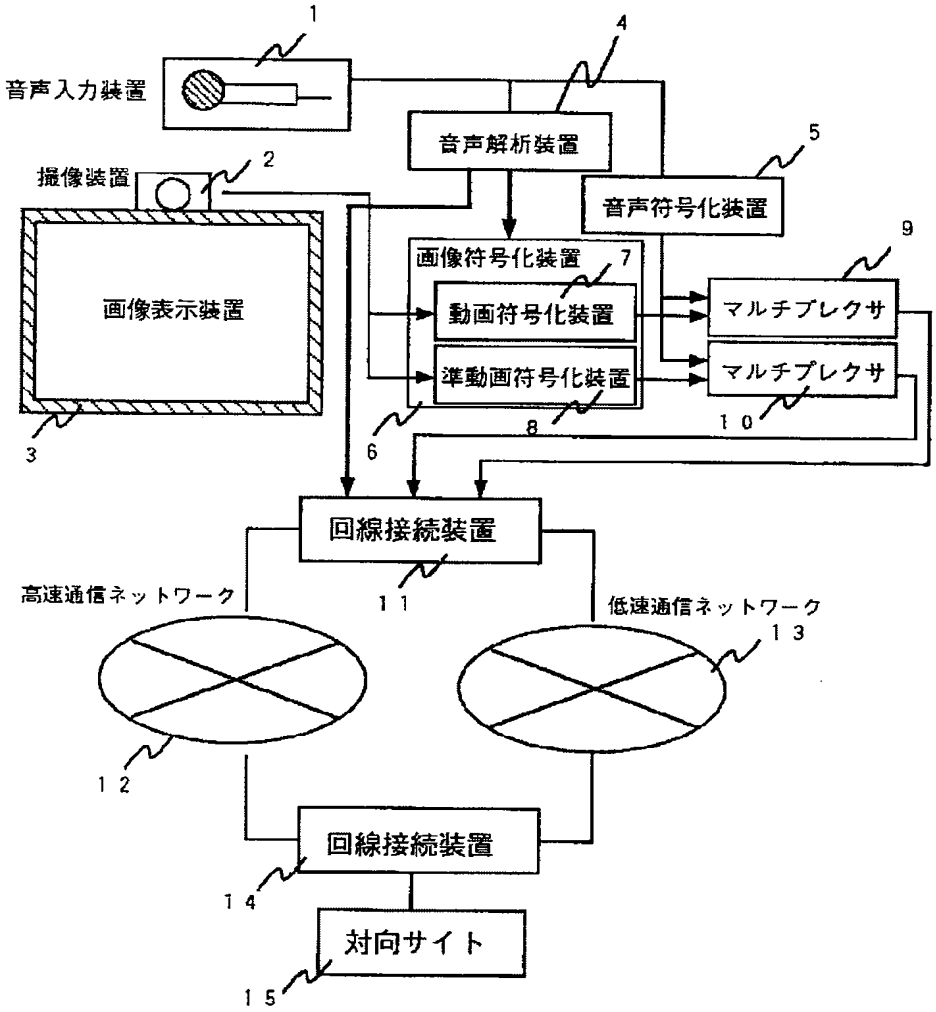
7

音声解析装置、５…音声符号化装置、６…画像符号化装置、７…動画符号化装置、８…準動画符号化装置、９、１０…マルチプレクサ、１１、１４…回線接続装置、１２…高速ネットワーク、１３…低速ネットワーク、１５…対向サイト。

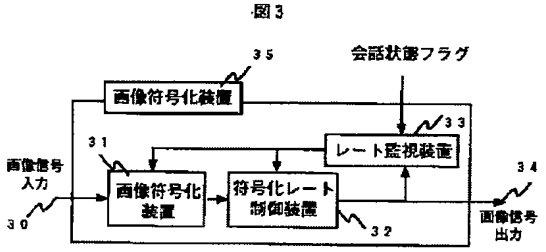
8

【図１】

図１

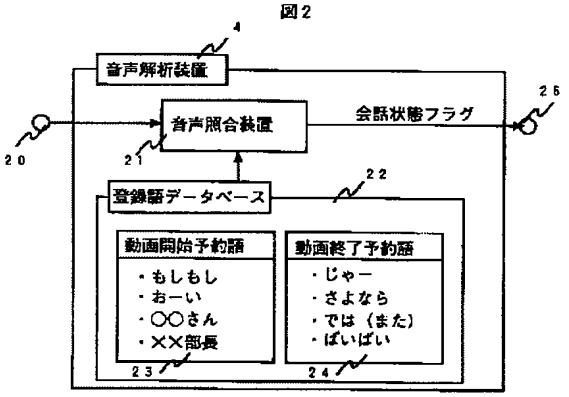


【図３】



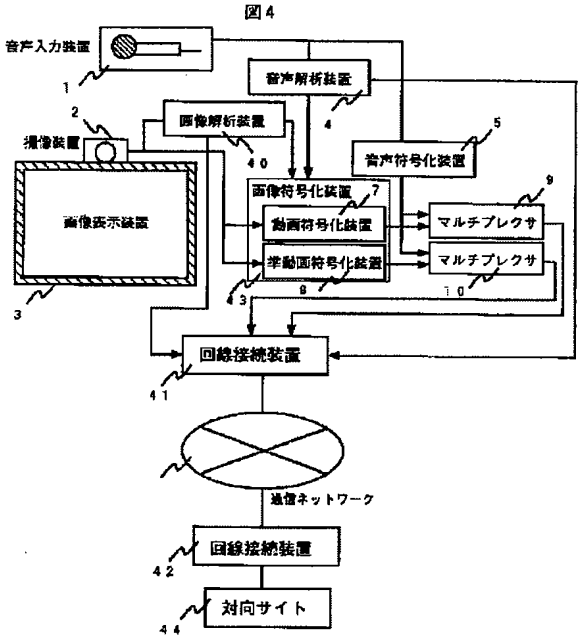
【図2】

図2



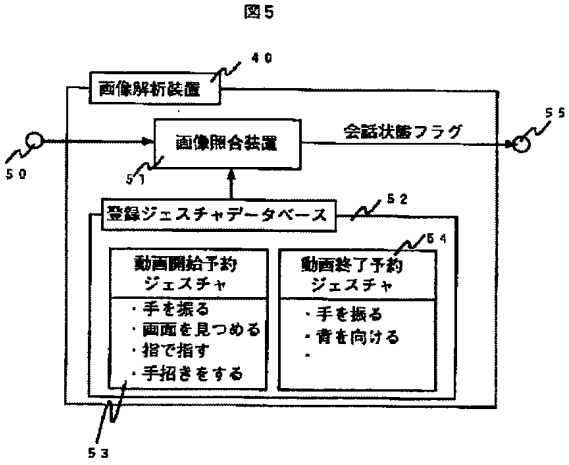
【図4】

図4



【図5】

図5



フロントページの続き

(72)発明者　木下　泰三
東京都国分寺市東恋ケ窪１丁目280番地
株式会社日立製作所中央研究所内